

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7700951号  
(P7700951)

(45)発行日 令和7年7月1日(2025.7.1)

(24)登録日 令和7年6月23日(2025.6.23)

(51)国際特許分類		F I			
G 0 6 T	1/00 (2006.01)	G 0 6 T	1/00	3 4 0 A	
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00	6 6 0 A	

請求項の数 6 (全15頁)

(21)出願番号	特願2024-502365(P2024-502365)	(73)特許権者	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(86)(22)出願日	令和4年2月25日(2022.2.25)	(74)代理人	110003708 弁理士法人鈴榮特許総合事務所
(86)国際出願番号	PCT/JP2022/007870	(72)発明者	蔵内 雄貴 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
(87)国際公開番号	WO2023/162132	(72)発明者	笹川 真奈 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
(87)国際公開日	令和5年8月31日(2023.8.31)	(72)発明者	萩山 直紀 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
審査請求日	令和6年6月4日(2024.6.4)	(72)発明者	佐野 文香

最終頁に続く

(54)【発明の名称】 画像変換装置、方法およびプログラム

(57)【特許請求の範囲】

【請求項1】

人の顔が含まれる画像から認識された顔パーツの特徴点を認識する特徴点認識部と、  
前記画像における顔が正面から認識できなくなる限界の角度に対する、正面からの前記顔の角度の比率と、前記顔の全体の領域に対する前記顔が物体で隠れている領域が除かれた領域の割合に基づいて、前記認識された顔の表情を変換すべき変換表情に変換するときの、前記変換表情に応じた前記顔パーツの特徴点のそれぞれについての変形量を表す変化量を補正する変化量補正部と、

前記補正した変化量により前記特徴点を変形することで前記人の顔の表情を変換した変換画像を得る表情変換部と、

を具備する、画像変換装置。

【請求項2】

前記変化量補正部は、

前記画像における顔が正面から認識できなくなる限界の角度に対する、正面からの前記顔の角度の比率と、前記顔の全体の面積に対する前記顔が物体で隠れている面積を除いた面積の割合を、前記顔パーツの特徴点のそれぞれについての所定の変化量に乗じることによって、前記変化量を補正する、

請求項1に記載の画像変換装置。

【請求項3】

前記顔が正面を向いているときの前記顔パーツの特徴点の3次元位置を回転または移動

したときの前記顔パーツの特徴点の2次元位置を算出し、前記算出した2次元位置と、現在の前記顔パーツの特徴点の2次元位置との距離の二乗和が最小になる角度を前記正面からの前記顔の角度として算出する、

請求項1に記載の画像変換装置。

【請求項4】

前記変換すべき変換表情ごとに、前記特徴点それぞれについての変形量を表す変化量が予め記憶される記憶装置と、

前記変換すべき変換表情を入力する変換表情入力部と、

を更に具備し、

前記変化量補正部は、

前記入力された変換表情に応じた前記変化量を前記記憶装置から読み出し、この読み出した変化量を補正する、

請求項1乃至3の何れか1項に記載の画像変換装置。

【請求項5】

人の顔の画像における表情を変換する画像変換装置により行われる方法であって、

前記画像変換装置の特徴点認識部により、人の顔が含まれる画像から認識された顔パーツの特徴点を認識することと、

前記画像変換装置の変化量補正部により、前記画像における顔が正面から認識できなくなる限界の角度に対する、正面からの前記顔の角度の比率と、前記顔の全体の領域に対する前記顔が物体で隠れている領域が除かれた領域の割合に基づいて、前記認識された顔の表情を変換すべき変換表情に変換するときの、前記変換表情に応じた前記顔パーツの特徴点のそれぞれについての変形量を表す変化量を補正することと、

前記画像変換装置の表情変換部により、前記補正した変化量により前記特徴点を変形することで前記人の顔の表情を変換した変換画像を得ることと、

を具備する画像変換方法。

【請求項6】

請求項1乃至4のいずれか1項に記載の画像変換装置の各部としてプロセッサを機能させる画像変換処理プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は、画像変換装置、方法およびプログラムに関する。

【背景技術】

【0002】

非特許文献1は、リアルタイムな表情変形（表情変換）フィードバックによる感情体験の操作の可能性について開示している。非特許文献1では、被験者の顔をリアルタイムにトラッキング（tracking）して自然な表情変形処理を施している。非特許文献1では、画像変換法としてRigid MLS（Moving Least Squares）法を使用して、顔画像における表情を変形している。Rigid MLS法は、画像から認識した画像中の特徴点を認識して、これを移動させることで、画像を歪めるという手法である。このような手法は非特許文献2にも開示される。なお、顔画像とは、被験者の顔を撮影した画像、コンピュータが生成したアバターの顔を抽出した画像、などである。

【先行技術文献】

【非特許文献】

【0003】

【文献】吉田成朗（Shigeo Yoshida）ら、「リアルタイムな表情変形フィードバックによる感情体験の操作（Manipulation of Emotional Experience by Real-time Deformed Facial Feedback）」、ヒューマンインタフェース学会論文誌（The Transactions of Human Interface Society）、Vol.17、No.1、2015

【文献】Tomas Jakab, et al., “Unsupervised Learning of Object Landmarks thro

10

20

30

40

50

ugh Conditional Image Generation ”, NIPS, 2018.

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、被験者の顔の角度が変わったり、顔の一部が隠れたりすることで、上記の特徴点の認識ができなかった場合、不自然なタイミング (timing) にて表情変換が止まってしまうため、不自然な変換による顔画像しか得ることができない。すなわち、顔の画像に表れる表情をシームレス (seamless) に変換することができない。

【0005】

この発明は、上記事情に着目してなされたもので、その目的とするところは、顔の画像に表れる表情をシームレスに変換することができるようにした画像変換装置、方法およびプログラムを提供することにある。

10

【課題を解決するための手段】

【0006】

上記課題を解決するために、この発明の一態様に係る画像変換装置は、人の顔が含まれる画像から認識された顔パーツの特徴点を認識する特徴点認識部と、前記画像における顔が正面から認識できなくなる限界の角度に対する、正面からの前記顔の角度の比率と、前記顔の全体の領域に対する前記顔が物体で隠れている領域が除かれた領域の割合に基づいて、前記認識された顔の表情を変換するべき変換表情に変換するときの、前記変換表情に応じた前記顔パーツの特徴点のそれぞれについての変形量を表す変化量を補正する変化量補正部と、前記補正した変化量により前記特徴点を変形することで前記人の顔の表情を変換した変換画像を得る表情変換部と、を備える。

20

【0007】

上記課題を解決するために、この一態様に係る画像変換方法は、人の顔の画像における表情を変換する画像変換装置により行われる方法であって、前記画像変換装置の特徴点認識部により、人の顔が含まれる画像から認識された顔パーツの特徴点を認識することと、前記画像変換装置の変化量補正部により、前記画像における顔が正面から認識できなくなる限界の角度に対する、正面からの前記顔の角度の比率と、前記顔の全体の領域に対する前記顔が物体で隠れている領域が除かれた領域の割合に基づいて、前記認識された顔の表情を変換するべき変換表情に変換するときの、前記変換表情に応じた前記顔パーツの特徴点のそれぞれについての変形量を表す変化量を補正することと、前記画像変換装置の表情変換部により、前記補正した変化量により前記特徴点を変形することで前記人の顔の表情を変換した変換画像を得ることと、を具備する。

30

【発明の効果】

【0008】

本発明によれば、顔の画像に表れる表情をシームレスに変換することができる。

【図面の簡単な説明】

【0009】

【図1】図1は、この発明の一実施形態に係る画像変換装置の構成の一例を示すブロック図 (block diagram) である。

40

【図2】図2は、画像変換装置のハードウェア (hardware) 構成の一例を示す図である。

【図3】図3は、顔の特徴点の一例を示す図である。

【図4】図4は、特徴点の記憶形態の一例を示す図である。

【図5】図5は、変化量の記憶形態の一例を示す図である。

【図6】図6は、画像変換装置による画像変換処理動作の一例を示すフローチャート (flow chart) である。

【図7】図7は、表示割合算出部により用いられるニューラルネットワーク (neural network) の一例を示す図である。

【図8】図8は、表示割合算出部により処理されるグリッドセル (grid cell) (グリッド領域) の一例を示す図である。

50

**【発明を実施するための形態】****【0010】****[一実施形態]**

以下、図面を参照して、この発明に係わる一実施形態を説明する。

**(構成例)**

図1は、この発明の一実施形態に係る画像変換装置の構成の一例を示すブロック図である。

図1に示される例では、この発明の一実施形態に係る画像変換装置100は、画像取得部11、特徴点認識部12、顔角度算出部13、表示割合算出部14、変換表情入力部15、変化量格納部16、変化量補正部17、表情変換部18、及び画像出力部19を有する。

10

**【0011】**

画像取得部11は、例えばwebカメラ(camera)により撮影された画像またはアバター(avatar)などからユーザ(user)の顔画像を取得する。画像取得部11は、取得した顔画像を、特徴点認識部12、表示割合算出部14、及び表情変換部18に出力する。

**【0012】**

特徴点認識部12は、画像取得部11が取得した顔画像を入力とし、その顔画像から認識される顔パーツ(parts)の特徴点を認識する。この特徴点認識部12における特徴点の認識手法については後述する。特徴点認識部12は、認識した特徴点を顔角度算出部13及び変化量補正部17に出力する。

20

**【0013】**

顔角度算出部13は、特徴点認識部12が認識した特徴点を入力とし、顔画像における顔の角度、例えば顔が正面を向いたときの位置を基準とした、顔の中心の現在の位置との間の角度(正面からの顔の角度と称することがある)を算出して、この算出した角度のデータ(data)を変化量補正部17に出力する。

**【0014】**

表示割合算出部14は、画像取得部11が取得した顔画像を入力とし、その顔画像に対して顔の全体のうち隠れている部分の割合を算出し、この算出した割合のデータを変化量補正部17に出力する。

**【0015】**

変換表情入力部15は、キーボード(keyboard)などのユーザインタフェース(user interface)からユーザが指定入力した、笑顔などの変換したい先の表情である変換表情(変換すべき変換表情と称することがある)を取得する。変換表情入力部15は、取得した変換表情を変化量補正部17に出力する。

30

**【0016】**

変化量格納部16には、変換したい先の表情ごとに、各特徴点についての変形量(座標値の移動量)を表す変化量が予め格納(記憶)される。変化量は、変換したい先の表情に応じて各特徴点の座標値を、どの程度移動すべきかを示す情報である。変化量は、例えば、ユーザが特定の顔画像について無表情顔に表情変形処理を適用しながら、自然な表情となるように調整して、予め求めることができる。

40

**【0017】**

変化量補正部17は、特徴点認識部12が認識した特徴点、顔角度算出部13により算出した顔角度、及び表示割合算出部14により算出した表示割合を入力する。

また、変化量補正部17は、変換表情入力部15から入力された変換表情で示される変換したい先の表情に応じた変化量を変化量格納部16から読み出す。

変化量補正部17は、これら入力した特徴点、顔角度、及び表示割合に基づいて、変換したい先の表情における変化量を後述する式によって補正した変化量を算出し、この算出した変化量のデータを表情変換部18に出力する。

**【0018】**

表情変換部18は、変化量補正部17が補正した変化量を入力とする。表情変換部18

50

は、上記補正した変化量、すなわち変換すべき変換表情に応じた変形量を表す変化量に基づいて、入力された顔画像における各特徴点を、入力した、その特徴点の補正した変化量である移動量に基づいて移動することで、顔画像の表情を変換した顔画像を得る。表情変換部 18 は、変換後の顔画像を画像出力部 19 に出力する。

#### 【0019】

画像出力部 19 は、表情変換部 18 からの変換後の顔画像を入力とし、入力された顔画像を出力する。ここで、出力とは、例えば、記憶媒体に記憶すること、ディスプレイ (display) で表示すること、通信ネットワークを介して他の機器へ送信すること、などを含む。

#### 【0020】

図 2 は、画像変換装置 100 のハードウェア構成の一例を示す図である。

画像変換装置 100 は、例えば、パーソナルコンピュータ (Personal computer)、スマートホン (smart phone)、サーバコンピュータ (server computer)、などのコンピュータにより構成される。画像変換装置 100 は、図 2 に示すように、CPU (Central Processing Unit) 等のハードウェアプロセッサ (hardware processor) (単にプロセッサと称することがある) 111A を有する。なお、CPU は、マルチコア (multi-core) 及びマルチスレッド (multithread) のものを用いることで、同時に複数の情報処理を実行することができる。また、プロセッサ 111A は、複数の CPU を備えていても良い。そして、画像変換装置 100 では、このプロセッサ 111A に対し、プログラムメモリ (program memory) 111B と、データメモリ (data memory) 112 と、通信

#### 【0021】

インタフェース 114 は、例えば一つ以上の有線または無線の通信モジュールを含むことができる。通信インタフェース 114 は、ケーブル (cable) もしくは LAN (Local Area Network) またはインターネット (internet) 等のネットワーク (NW) を介して接続される他のコンピュータおよび web カメラ、などとの間で通信を行うことができる。

#### 【0022】

入出力インタフェース 113 には、入力デバイス (device) 200 及び出力デバイス 300 が接続されている。入力デバイス 200 は、キーボード、マウス (mouse) などのポインティングデバイス (pointing device)、などの入力デバイス、カメラなどのセンサデバイス (sensor device)、などを含む。また、出力デバイス 300 は、液晶ディスプレイ、CRT (Cathode Ray Tube) ディスプレイ、などの表示デバイスである。入力デバイス 200 及び出力デバイス 300 は、いわゆるタブレット (tablet) 型の入力・表示デバイスを用いたものが用いられることもできる。この種の入力・表示デバイスは、例えば液晶または有機 EL (Electro Luminescence) を使用した表示デバイスの表示画面上に、静電方式または圧力方式を採用した入力検知シート (sheet) を配置して構成される。入出力インタフェース 113 は、上記入力デバイス 200 において入力された操作情報をプロセッサ 111A に入力すると共に、プロセッサ 111A で生成された表示情報を出力デバイス 300 に表示させる。

#### 【0023】

なお、入力デバイス 200 及び出力デバイス 300 は、入出力インタフェース 113 に接続されていなくても良い。入力デバイス 200 及び出力デバイス 300 は、通信インタフェース 114 と直接またはネットワークを介して接続するための通信ユニットを備えることで、プロセッサ 111A との間で情報の授受を行い得る。

#### 【0024】

また、入出力インタフェース 113 は、フラッシュメモリ (Flash memory) 等の半導体メモリといった記録媒体のリード/ライト (read / write) 機能を有しても良いし、あるいは、そのような記録媒体のリード/ライト機能を持ったリーダライタ (reader write

10

20

30

40

50

r)との接続機能を有しても良い。さらに、入出力インタフェース113は、他の機器との接続機能を有して良い。

#### 【0025】

プログラムメモリ111Bは、非一時的な有形のコンピュータ可読記憶媒体として、随時書込み及び読出しが可能な不揮発性メモリ(non-volatile memory)と、随時読出しのみが可能な不揮発性メモリとが組み合わせて使用されたものである。随時書込み及び読出しが可能な不揮発性メモリは、例えば、HDD(Hard Disk Drive)、SSD(Solid State Drive)、などである。随時読出しのみが可能な不揮発性メモリは、例えば、ROM(Read Only Memory)などである。このプログラムメモリ111Bには、プロセッサ111Aが一実施形態に係る各種制御処理を実行するために必要なプログラム、例えば画像変換プログラムが格納されている。すなわち、上記の画像取得部11、特徴点認識部12、顔角度算出部13、表示割合算出部14、変換表情入力部15、変化量補正部17、表情変換部18、及び画像出力部19の各部における処理機能部は、何れも、プログラムメモリ111Bに格納された画像変換プログラムを上記プロセッサ111Aにより読み出させて実行させることにより実現され得る。なお、これらの処理機能部の一部または全部は、特定用途向け集積回路(ASIC: Application Specific Integrated Circuit)またはFPGA(field-programmable gate array)等の集積回路を含む、他の多様な形式によって実現されても良い。

10

#### 【0026】

データメモリ112は、有形のコンピュータ可読記憶媒体として、例えば、上記の不揮発性メモリと、RAM(Random Access Memory)等の揮発性メモリ(volatile memory)とが組み合わせて使用されたものである。このデータメモリ112は、各種処理が行われる過程で取得及び作成された各種データが記憶されるために用いられる。すなわち、データメモリ112には、各種処理が行われる過程で、適宜、各種データを記憶するための領域が確保される。

20

#### 【0027】

図3は、顔の特徴点の一例を示す図である。図3中の星印が、プロセッサ111Aが認識した特徴点であり、各特徴点の横に付された数字は各特徴点を識別するための一意な特徴点ID(Identifier)である。特徴点IDの数及び各特徴点IDに対する顔の部分は、採用する特徴点認識手法により決まっている。例えば、特徴点ID「18」の特徴点は向かって左の眉の左端、のように予め決まっている。

30

#### 【0028】

図4は、特徴点の記憶形態の一例を示す図である。図4に示すように、データメモリ112には、テーブル(table)形式で、特徴点IDに対応付けて顔画像中の特徴点のx座標及びy座標が記憶される。座標の値はピクセル(pixel)である。従って、データメモリ112には、図3の例であれば、特徴点ID「1」~「68」に係る特徴点について、そのxy座標が記憶される。

#### 【0029】

データメモリ112には、プロセッサ111Aが上記の変換表情入力部15として動作したときに取得した、ユーザによって指定された変換表情が記憶される。

40

データメモリ112には、上記の変化量格納部16に格納される変換量が格納され得る。

#### 【0030】

図5は、変化量の記憶形態の一例を示す図である。図5に示すように、データメモリ112には、変換表情ごとに、特徴点IDに対応付けて、特徴点のx座標の変化量とy座標の変化量とが、被写体である人物によらない変化量として、テーブル形式で記憶される。変化量の値はピクセルである。変化量は、特徴点の移動方向と移動量によって表される。例えば、移動量「+1」は、正方向に1ピクセル移動することを表す。

#### 【0031】

データメモリ112には、プロセッサ111Aが上記の表情変換部18として動作したときに変換した顔画像が記憶され得る。

50

また、データメモリ 112 には、プロセッサ 111A が動作途中で発生する種々の中間データが記憶され得る。

【0032】

(動作)

次に、画像変換装置 100 の動作を説明する。

図 6 は、画像変換装置 100 による画像変換処理動作の一例を示すフローチャートである。画像変換装置 100 のプロセッサ 111A は、プログラムメモリ 111B に記憶された画像変換プログラムを読み出して実行することで、このフローチャートに示す画像変換装置 100 としての動作を開始する。プロセッサ 111A での画像変換プログラムの実行は、入力デバイス 200 から、入出力インタフェース 113 を介して、あるいは、通信インタフェース 114 を介して、画像変換の実施を指示されることで開始される。

10

【0033】

プロセッサ 111A は、変換表情入力部 15 として動作して、ユーザによる、笑顔などの変換したい先の表情である変換表情の指定入力待 (ステップ S1)。例えば、プロセッサ 111A は、入出力インタフェース 113 または通信インタフェース 114 を介した入力デバイス 200 からの入力信号が変換表情の指定入力を含むか否かを判断する。変換表情の指定入力があったならば、プロセッサ 111A は、ステップ S2 の処理へ移行する。

【0034】

プロセッサ 111A は、指定された変換表情を、データメモリ 112 に記憶させる (ステップ S2)。

20

【0035】

プロセッサ 111A は、画像取得部 11 として動作して、顔画像を取得する (ステップ S3)。例えば、プロセッサ 111A は、入力デバイス 200 のカメラによる被験者の顔の撮影画像を入出力インタフェース 113 を介して取得する。あるいは、プロセッサ 111A は、ネットワークに接続された web カメラにより撮影された顔画像または他のコンピュータが生成したアバターの顔を通信インタフェース 114 を介して取得する。プロセッサ 111A は、取得した顔画像を、データメモリ 112 に記憶させる。

【0036】

プロセッサ 111A は、特徴点認識部 12 として動作して、データメモリ 112 に記憶されている顔画像から特徴点を認識する (ステップ S4)。プロセッサ 111A は、例えば、`dlib` の `face_landmark_detection` 関数 (例えば [http://dlib.net/face\\_landmark\\_detection.py.html](http://dlib.net/face_landmark_detection.py.html) を参照) などを利用して、顔画像に対して特徴点を認識する。具体的には、プロセッサ 111A は、入力の顔画像に対して、HOG (Histogram of Oriented Gradients) 特徴と呼ばれる輝度の勾配方向の分布を抽出する。HOG 特徴と顔の特徴点の位置を紐付けたデータをもとに学習されたモデル (model) は一般的に提供されている。よって、プロセッサ 111A は、抽出された HOG 特徴を、この学習モデルに入力し、顔の特徴点の位置を取得する。プロセッサ 111A は、取得した特徴点の位置をデータメモリ 112 に記憶させる。

30

【0037】

プロセッサ 111A は、顔角度算出部 13 として動作して、例えば `opencv` などを利用して、顔画像における顔の角度を算出する (ステップ S5)。

40

具体的には、プロセッサ 111A は、顔が正面に向いているときの顔パーツの特徴点の 3次元位置 ( $P_{3d}$ ) を予め計測して、これをデータメモリ 112 に保持する。

プロセッサ 111A は、顔画像の顔パーツの現在の特徴点の 2次元位置 ( $P'_{2d}$ ) を取得する。

プロセッサ 111A は、上記 3次元位置 ( $P_{3d}$ ) を回転または移動したときの顔パーツの特徴点の 2次元位置 ( $P_{2d}$ ) を算出する。

プロセッサ 111A は、例えば `opencv` の `ProjectPoints2` 関数 (例えば [http://opencv.jp/opencv-2svn/py/camera\\_calibration\\_and\\_3d\\_reconstruction.html#projectp](http://opencv.jp/opencv-2svn/py/camera_calibration_and_3d_reconstruction.html#projectp)

50

oints2を参照)などを利用して、上記各2次元位置を算出する。

【0038】

プロセッサ111Aは、2次元位置(P<sub>2d</sub>)と、2次元位置(P'<sub>2d</sub>)の距離の二乗和(sum of squares)(D)を算出する。

プロセッサ111Aは、この二乗和Dを最小化するような角度(および移動量)を大域的最適化(global optimization)により求める。

【0039】

プロセッサ111Aは、例えばopencvのsolvePnP関数(例えば[http://opencv.jp/opencv-2svn/cpp/camera\\_calibration\\_and\\_3d\\_reconstruction.html#cv\\_solvepnp](http://opencv.jp/opencv-2svn/cpp/camera_calibration_and_3d_reconstruction.html#cv_solvepnp)を参照)などを利用して、上記最小化するような角度(および移動量)を正面からの顔の角度(a)として算出部する。

10

【0040】

プロセッサ111Aは、顔認識ツールを起動しつつ顔を動かしながら、認識ができなくなった際の特徴点の位置を取得することにより、認識ができる限界の顔の角度(A)を被写体の人物によらない角度として予め算出し、これをデータメモリ112に保持する。

【0041】

次に、プロセッサ111Aは、表示割合算出部14として動作して、顔画像に対して顔の全体の領域のうち顔以外の物体で隠れている領域の割合である、顔の表示割合を算出する(ステップS6)。例えば顔の全体の10%が顔以外の物体で隠れていれば、上記顔の表示割合は10%となる。

20

【0042】

ここで、表示割合算出部14による算出の例を図7および図8を参照して説明する。

【0043】

図7は、表示割合算出部により用いられるニューラルネットワークの一例を示す図である。図8は、表示割合算出部により処理されるグリッドセルの一例を示す図である。ここでは、動物および各種物体が含まれる入力画像に係る例を説明するが、これらが人の顔および顔を隠している物体、例えば手またはその他の物体であるときにも同様に適用が可能である。

【0044】

図7および図8に示された例では、既知のYOLO(You Only Look Once)(ディープラーニング(deep learning)による一般物体検出手法)が用いられ得る。この手法は、例えば下記の資料に開示される。

30

「Joseph Redmon, et al., "YOLOv3: An Incremental Improvement", arXiv preprint, arXiv:1804.02767, 2018.」

【0045】

この手法では、プロセッサ111Aは、顔画像を正方形にリサイズ(resize)し、これを図7に示されるような、画像処理の分野で数多く用いられるニューラルネットワークであるCNN(Convolutional Neural Network(畳み込みニューラルネットワーク))に入力する。プロセッサ111Aは、図7に示されたCNNにおける24層の畳み込み層(Conv. Layer)および4層のpooling層(図7の符号a参照)を経て顔画像から特徴を抽出し、2層の全結合層(Conn. Layer)で(図7の符号b参照)、画像における物体のBounding Box、および物体の種類の確率を推定することができる。畳み込み層の最終出力サイズ7×7はgrid cellの分割数と一致する。

40

【0046】

上記入力された画像は、図8に示されるような、S×Sのgrid cellに分割される(図8の(a)参照)。

プロセッサ111Aは、上記分割した各grid cellに対して、B個の物体のBounding Boxを推定する。プロセッサ111Aは、1つのBounding Boxにつき、Bounding Boxの座標値、幅、高さ(x, y, w, h)と、そのBounding Boxが物体である信頼度(confidence)スコアでなる、計5つの値を出力する(図8の(b)参照)。

50

## 【 0 0 4 7 】

座標値のx, yは、grid cellの境界を基準にしたBounding Boxの中心座標であり、幅wと高さhは画像全体のサイズに対する相対値であり、信頼度スコア (score) は、そのBounding Boxが物体か背景かの確率を表す。この確率は、物体なら「1」で背景であれば「0」である。

## 【 0 0 4 8 】

物体領域の推定精度を測る指標として、正解Bounding Boxと推定Bounding Boxの一致具合を表すIoU (Intersection over Union)がある。上記YOLOではBounding Boxの信頼度スコアがIoUを表す。

## 【 0 0 4 9 】

プロセッサ111Aは、各grid cell単位で物体の種類を推定する。例えば、プロセッサ111Aは、C種類の分類クラス (classification class) で、grid cellが物体である場合に、どのクラスに属するかの確率、すなわち条件付き確率 (conditional probability) を推定する (図8の(c)参照)。

## 【 0 0 5 0 】

プロセッサ111Aは、ここで推定したクラス確率を上記のBounding Boxと統合することで、何の物体であるかを示す複数のBounding Boxを得る (図8の(d)参照)。

## 【 0 0 5 1 】

プロセッサ111Aは、重複領域も含んだ、これらのBounding Boxを、信頼度スコアの高いBounding Boxを基準にNMS (Non-Maximum Suppression) という手法で選別する (図8の(e)参照)。NMSは、IoU値が大きい(重なり度合いの高い)領域をしきい値で抑制 (suppression) する。これにより物体領域の検出結果が得られる。

## 【 0 0 5 2 】

プロセッサ111Aは、顔領域と、この領域に重畳する物体領域があったときは、重畳している領域の面積を顔領域の面積で除することによって、上記の顔の表示割合を算出することができる。

## 【 0 0 5 3 】

次に、プロセッサ111Aは、変化量補正部17として動作して、変換したい先の表情に応じた変化量を変化量格納部16から読み出し、S4で認識した特徴点、S5で算出した顔角度、及びS6で算出した表示割合に基づいて、変換したい先の表情に応じた、上記読み出した変化量を補正した変化量を算出する (ステップS7)。

## 【 0 0 5 4 】

具体的には、プロセッサ111Aは、顔の角度、すなわち正面からの顔の角度aおよび認識ができる限界の顔の角度Aと、顔全体の領域に対する顔が隠れている領域の割合Hを取得し、これらに応じて、下記の式(1)により、表情変換の変化量を減衰させる、すなわち変化量を補正し、この補正した結果をデータメモリ112に保持する。

$$P_{new} = P \cdot (1 - H) \cdot a / \dots \text{式(1)}$$

式(1)の左辺  $P_{new}$  は、表情変換の減衰させた、すなわち補正後の変化量であり、右辺の  $P$  は表情変換の補正前の変化量である。

## 【 0 0 5 5 】

すなわち、上記の例では、(1) 正面からの顔の角度aおよび認識ができる限界の顔の角度Aとの比率  $a /$  と、(2) 顔全体の領域に対する顔が隠れている領域の割合Hと、に基づいて、補正後の変化量が算出される。

なお、この例に限らず、例えば、許容される精度の範囲内で、(1) 正面からの顔の角度aおよび認識ができる限界の顔の角度Aとの比率  $a /$  と、(2) 顔全体の領域に対する顔が隠れている領域の割合Hと、の一方に基づいて補正後の変化量が算出されてもよい。

## 【 0 0 5 6 】

このようにして変化量を補正すれば、顔の角度が変わったり、顔の一部が隠れたりすることにより、特徴点の認識ができなかったとしても、自然でないタイミングで表情変換が止まることなく、顔画像の表情を自然に変換することができる。

10

20

30

40

50

## 【 0 0 5 7 】

プロセッサ 1 1 1 A は、表情変換部 1 8 として動作して、データメモリ 1 1 2 に記憶されている顔画像の表情を変換する（ステップ S 8）。すなわち、プロセッサ 1 1 1 A は、データメモリ 1 1 2 に記憶された、変換表情に応じた変化量が補正された結果に基づいて、顔画像を変換する。例えば、プロセッサ 1 1 1 A は、M L S の実装（例えば <https://github.com/Jarvis73/Moving-Least-Squares> を参照）などを利用する。

## 【 0 0 5 8 】

具体的には、プロセッサ 1 1 1 A は、各特徴点について、データメモリ 1 1 2 に記憶された変換表情に応じた変化量の補正後の変化量分だけ移動させる。例えば、表情を笑顔に変換する場合には、特徴点 I D 「 1 」 の制御点については、変換前の x y 座標が ( 2 3 , 4 5 ) であるので（図 4 参照）、プロセッサ 1 1 1 A は、x 座標を「 + 1」、y 座標を「 + 2」する（図 5 参照）ことで、当該特徴点の画素を ( 2 4 , 4 7 ) に移動するような変換を行う。

10

## 【 0 0 5 9 】

そして、特徴点については、プロセッサ 1 1 1 A は、下記の式 ( 2 ) に示されるアフィン ( Affine ) 変換 ( ヘルムート ( Helmert ) 変換 = 相似変換及び r i g i d d e f o r m a t i o n = 剛体変形を含む ) を適用する。

## 【 0 0 6 0 】

## 【 数 1 】

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad \dots \text{式(2)}$$

20

## 【 0 0 6 1 】

ただし、上記式 ( 2 ) の x , y は近傍の特徴点の座標であり、x' , y' は、その特徴点の座標に変化量を足した座標であり、a , b , c , d はパラメータ ( parameter ) であり、t<sub>x</sub> , t<sub>y</sub> は平行移動パラメータである。プロセッサ 1 1 1 A は、特徴点の座標 x , y と変化量を足した座標 x' , y' の最小二乗平均 ( least square means ) を算出し、これを最小化するようなパラメータ a , b , c , d , t<sub>x</sub> , t<sub>y</sub> を大域的最適化により求める。そして、プロセッサ 1 1 1 A 変換するべき対象点の座標を x , y として、これら求めたパラメータを用いて変換後の座標を求める。プロセッサ 1 1 1 A は、こうして求めたパラメータ a , b , c , d , t<sub>x</sub> , t<sub>y</sub> を用いて、特徴点から上記アフィン変換により変換した後の座標を求める。

30

## 【 0 0 6 2 】

プロセッサ 1 1 1 A は、こうして変換した後の顔画像を変換画像としてデータメモリ 1 1 2 に記憶させる。

## 【 0 0 6 3 】

プロセッサ 1 1 1 A は、画像出力部 1 9 として動作して、データメモリ 1 1 2 に記憶された変換画像を出力する（ステップ S 9）。例えば、プロセッサ 1 1 1 A は、入出力インタフェース 1 1 3 を介して出力デバイス 3 0 0 に顔画像を表示させる。あるいは、プロセッサ 1 1 1 A は、通信インタフェース 1 1 4 によりネットワーク上に送信し、ネットワークに接続された表示デバイスに表示させたり、ネットワークに接続された他のコンピュータの表示部に表示させたりする。

40

## 【 0 0 6 4 】

プロセッサ 1 1 1 A は、図 6 のフローチャートに示す画像変換装置 1 0 0 としての動作を終了するか否か判断する（ステップ S 1 0）。例えば、プロセッサ 1 1 1 A は、入力デバイス 2 0 0 から、入出力インタフェース 1 1 3 を介して、あるいは、通信インタフェース 1 1 4 を介して、ユーザから画像変換の終了を指示されたか否か確認する。ここで、上記動作を終了する場合には（ステップ S 1 0 の Y E S）、プロセッサ 1 1 1 A は、図 6 のフローチャートに示す動作を終了する。

50

## 【 0 0 6 5 】

これに対して、未だ上記動作を終了しない場合には（ステップ S 1 0 の N O ）、プロセッサ 1 1 1 A は、変換表情入力部 1 5 として動作して、ユーザによる変換表情の変更指定入力が有ったか否かを判断する（ステップ S 1 1 ）。変換表情の変更指定入力が無ければ（ステップ S 1 1 の N O ）、プロセッサ 1 1 1 A は、ステップ S 3 の処理へ移行する。また、変換表情の変更指定入力が有った場合には（ステップ S 1 0 の Y E S ）、プロセッサ 1 1 1 A は、ステップ S 2 の処理へ移行する。

## 【 0 0 6 6 】

以上に説明した一実施形態に係る画像変換装置 1 0 0 は、顔角度算出部 1 3 と、表示割合算出部 1 4 と、変化量補正部 1 7 と、表情変換部 1 8 とを備える。表情変換部 1 8 は、  
10 変換すべき変換表情に応じた変形量により特徴点を変換することで人の顔の表情を変換した変換画像を得る。

従って、一実施形態に係る画像変換装置 1 0 0 は、顔の角度が変わったり、顔の一部が隠れたりすることにより、特徴点の認識ができなかったとしても、自然でないタイミングで表情変換が止まることなく、顔画像の表情を自然に変換することができる。

## 【 0 0 6 7 】

[ 他の実施形態 ]

なお、この発明は上記一実施形態に限定されるものではない。

例えば、以上で説明した各処理の流れは、説明した手順に限定されるものではなく、いくつかのステップの順序が入れ替えられても良いし、いくつかのステップが同時並行で実施されても良い。  
20

## 【 0 0 6 8 】

また、以上で説明した各処理の流れは、リアルタイムに取得する顔画像の表情をリアルタイムに変換していく場合であったが、リアルタイム処理ではなく、保存された顔画像の表情を変換する用途にも同様に適用できる。

## 【 0 0 6 9 】

また、各実施形態に記載された手法は、計算機（コンピュータ）に実行させることができるプログラム（ソフトウェア手段）として、例えば磁気ディスク（フロッピー（登録商標）ディスク（Floppy disk）、ハードディスク（hard disk）等）、光ディスク（optical disc）（CD-ROM、DVD、MO等）、半導体メモリ（ROM、RAM、フラッシュメモリ等）等の記録媒体に格納し、また通信媒体により伝送して頒布され得る。なお、媒体側に格納されるプログラムには、計算機に実行させるソフトウェア手段（実行プログラムのみならずテーブル、データ構造も含む）を計算機内に構成させる設定プログラムをも含む。本装置を実現する計算機は、記録媒体に記録されたプログラムを読み込み、また場合により設定プログラムによりソフトウェア手段を構築し、このソフトウェア手段によって動作が制御されることにより上述した処理を実行する。なお、本明細書でいう記録媒体は、頒布用に限らず、計算機内部あるいはネットワークを介して接続される機器に設けられた磁気ディスク、半導体メモリ等の記憶媒体を含むものである。  
30

## 【 0 0 7 0 】

なお、本発明は、上記実施形態に限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で種々に変形することが可能である。また、各実施形態は適宜組み合わせ実施してもよく、その場合組み合わせた効果が得られる。更に、上記実施形態には種々の発明が含まれており、開示される複数の構成要件から選択された組み合わせにより種々の発明が抽出され得る。例えば、実施形態に示される全構成要件からいくつかの構成要件が削除されても、課題が解決でき、効果が得られる場合には、この構成要件が削除された構成が発明として抽出され得る。  
40

## 【 符号の説明 】

## 【 0 0 7 1 】

1 0 0 ... 画像変換装置

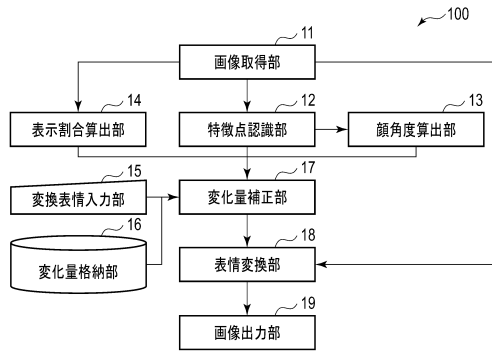
1 1 ... 画像取得部  
50

- 1 2 ... 特徴点認識部
- 1 3 ... 顔角度算出部
- 1 4 ... 表示割合算出部
- 1 5 ... 変換表情入力部
- 1 6 ... 変化量格納部
- 1 7 ... 変化量補正部
- 1 8 ... 表情変換部
- 1 9 ... 画像出力部
- 1 1 1 A ... プロセッサ
- 1 1 1 B ... プログラムメモリ
- 1 1 2 ... データメモリ
- 1 1 3 ... 入出力インタフェース
- 1 1 4 ... 通信インタフェース
- 1 1 5 ... バス
- 2 0 0 ... 入力デバイス
- 3 0 0 ... 出力デバイス

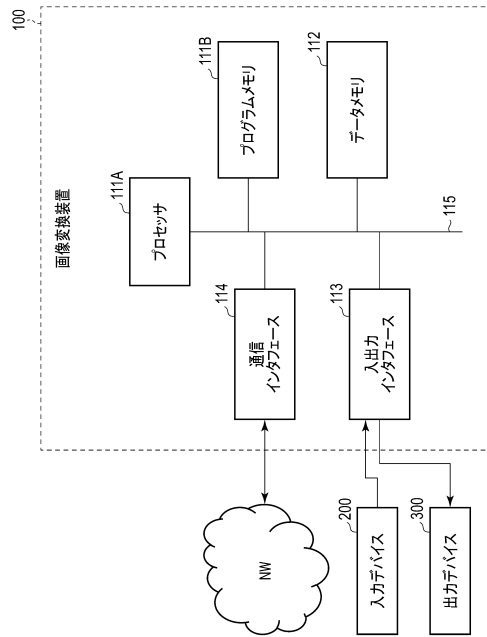
10

【図面】

【図 1】



【図 2】



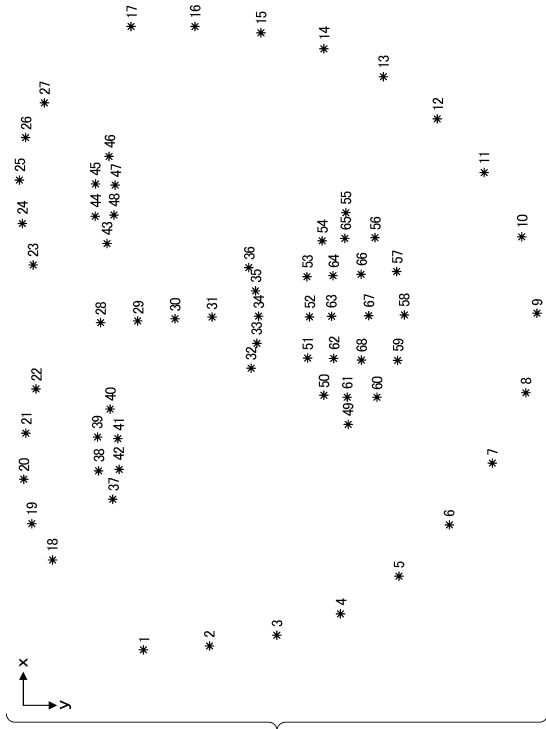
20

30

40

50

【図 3】



【図 4】

特徴点ID	x座標	y座標
1	23	45
2	67	89
⋮	⋮	⋮

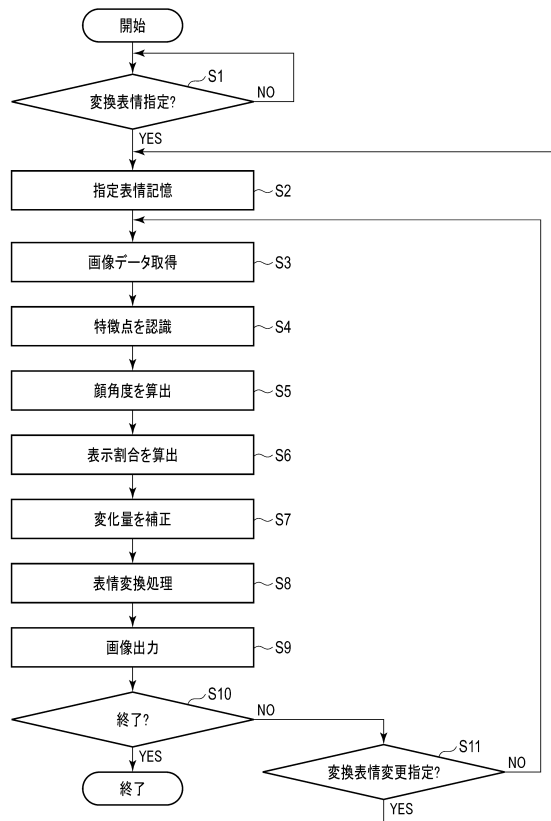
10

20

【図 5】

変換表情	特徴点ID	x座標	y座標
笑顔	1	+1	+2
	2	+5	-6
⋮	⋮	⋮	⋮
悲しみ	1	-3	-4
	2	-7	+8
⋮	⋮	⋮	⋮

【図 6】

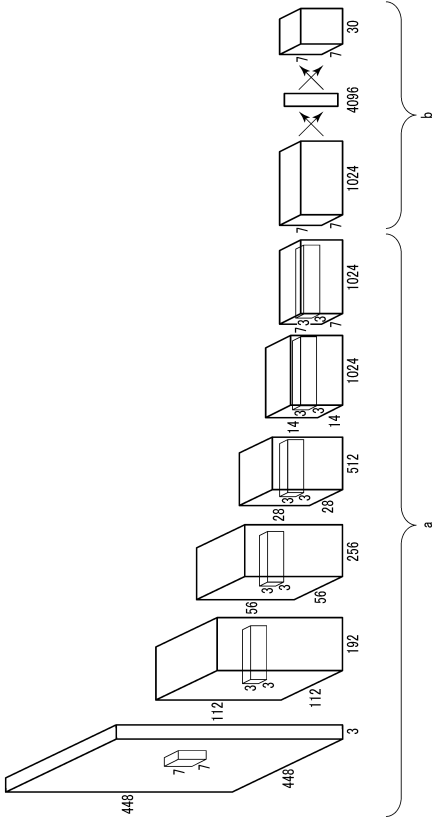


30

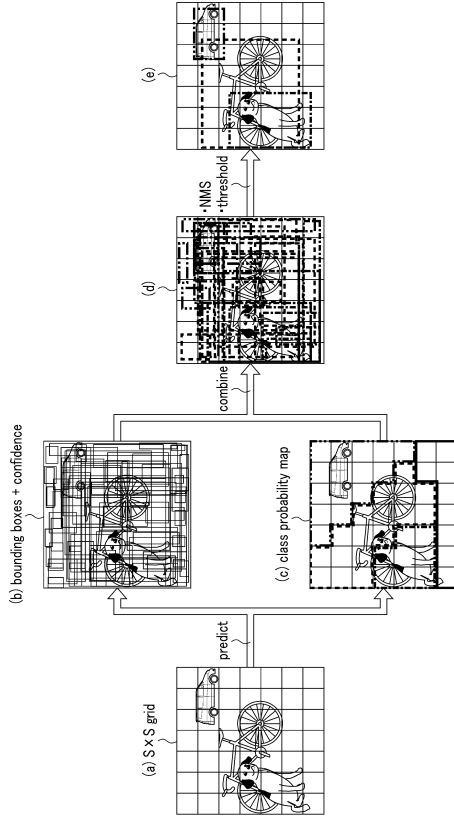
40

50

【 図 7 】



【 図 8 】



10

20

30

40

50

## フロントページの続き

- 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
- (72)発明者 山本 隆二  
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
- 審査官 鈴木 圭一郎
- (56)参考文献 特開2011-060038(JP,A)  
特開2005-215763(JP,A)  
特開2021-077376(JP,A)  
中国特許出願公開第112766063(CN,A)
- (58)調査した分野 (Int.Cl., DB名)  
G06T 1/00  
G06T 7/00